Linde A.C. De Grande, Kenneth Goossens, Katleen Van Uytfanghe, Dietmar Stöckl and Linda M. Thienpont*

The Empower project – a new way of assessing and monitoring test comparability and stability

DOI 10.1515/cclm-2014-0959
Received September 29, 2014; accepted December 15, 2014

Abstract

Background: Manufacturers and laboratories might benefit from using a modern integrated tool for quality management/assurance. The tool should not be confounded by commutability issues and focus on the intrinsic analytical quality and comparability of assays as performed in routine laboratories. In addition, it should enable monitoring of long-term stability of performance, with the possibility to quasi "real-time" remedial action. Therefore, we developed the "Empower" project.
Methods: The project comprises four pillars: (i) master comparisons with panels of frozen single-donation samples, (ii) monitoring of patient percentiles and (iii) internal quality control data, and (iv) conceptual and statistical education about analytical quality. In the pillars described here (i and ii), state-of-the-art as well as biologically derived specifications are used.
Results: In the 2014 master comparisons survey, 125 laboratories forming 8 peer groups participated. It showed not only good intrinsic analytical quality of assays but also assay biases/non-comparability. Although laboratory performance was mostly satisfactory, sometimes huge between-laboratory differences were observed. In patient percentile monitoring, currently, 100 laboratories participate with 182 devices. Particularly, laboratories with a high daily throughput and low patient population variation show a stable moving median in time with good between-instrument concordance. Shifts/drifts due to lot changes are sometimes revealed. There is evidence that outpatient medians mirror the calibration set-points shown in the master comparisons.
Conclusions: The Empower project gives manufacturers and laboratories a realistic view on assay quality/comparability as well as stability of performance and/or the reasons for increased variation. Therefore, it is a modern tool for quality management/assurance toward improved patient care.

Keywords: analytical/population variation; bias; drift; median; moving median; outpatients; quality indicators; shift.

Introduction

Manufacturers and laboratories have common interest in precise, unbiased, and stable in vitro diagnostic assays enabling optimal patient care. Although they both monitor the above test attributes, they have different objectives and access to existing data, which are facts that might hamper the dialogue between them. For example, manufacturers are mainly interested in the global performance of their assays (=peer performance), whereas laboratories rather focus on their own performance. However, for troubleshooting purposes, peer performance is also of interest to laboratories. Manufacturers monitor laboratories by an online link with their systems, whereas laboratories have easy access to their own data. The data sources can be bridged by independent third-party programs for peer group-based combined internal quality control (IQC)/external quality assessment (EQA). However, this approach has limitations. Commutability issues of the used materials make that peer group assessment cannot give information on trueness of performance. Additionally, it may cause that variations in patient data (e.g., trends and shifts due to reagent lot changes) are not well reflected [1–3]. Besides, continuous monitoring of the results is rather the exception, and even if done, the data are usually not accessible in real time. In addition, the external program providers mostly do not critically review or publish the data, but

***Corresponding author: Linda M. Thienpont,** Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium, Phone: +32-9-264-8104, Fax: +32-9-264-8198, E-mail: linda.thienpont@ugent.be
Linde A.C. De Grande, Kenneth Goossens and Katleen Van Uytfanghe: Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium
Dietmar Stöckl: STT-Consulting, Horebeke, Belgium

leave the interpretation to the participating laboratories. This practice is of course driven by the commercial surrounding in which they operate, which hampers them to disclose performance data of individual diagnostic manufacturers. In contrast, independent national or regional EQA schemes theoretically are in the position to openly demonstrate the performance of commercial test systems. However, this requires that sufficient laboratories participate, so that the peer groups can be well defined. This is for most of the schemes not possible, therefore, they rather restrict to assessing the competence of laboratories [2, 3]. This is in turn limited by the fact that EQA schemes seldom work with fully commutable materials, conduct surveys at low frequency and report retrospectively. From this perspective, it would be desirable to implement an independently operated "online" tool that enables to monitor comparability and stability between peer groups and laboratories without being confounded by non-commutability issues because it uses real patient samples. To maximize the utility of the tool, the information should be shared between participants and manufacturers but within confidentiality constraints. This means that an individual evaluation report should only be available to the laboratory to which it applies. The tool could in the same time serve to empower laboratories for the future tasks they face, among others, providing input for the development and implementation of global health-care policies.

In response to these needs, we initiated the so-called Empower project. It is intended to establish a bottom-up cooperation between laboratories and manufacturers, so that they can pursue the common objective of assessing and improving test comparability and stability, whereby we see our role as independent third party mediator. It is our strong belief that such a transparent cooperation will be of benefit to all stakeholders involved in laboratory medicine. The project stands on four pillars: (i) master comparisons with fresh-frozen single-donation serum samples; (ii) monitoring of patient percentiles and (iii) IQC, both across laboratories and manufacturers; (iv) conceptual and statistical education about analytical quality in the medical laboratory (e.g., analytical performance specifications) and elaboration of statistically sound and "actionable" experiments for analytical quality management and assurance. Laboratories are free to participate in all pillars of our project or to select the most appropriate one(s) for their purpose. Here, we report on the status of the project with respect to the master comparisons and patient percentile monitoring and share the first observations on test comparability and stability of performance.

Materials and methods

Master comparisons

As previously described, we conduct the master comparison surveys for diagnostic assays with panels of 20 fresh-frozen, single-donation, commutable serum samples (each available in a volume of approx. 180 mL) [4–6]. The samples are prepared using the Clinical and Laboratory Standards Institute C37-A protocol, however, without pooling and filtration and are dispatched on dry ice [7]. Participation is made conditional of the use of a homogeneous test system, i.e., instrument, reagent, and calibrator from the same manufacturer. The number and selection of laboratories is adapted to obtain peer groups representing the main manufacturers/diagnostic test systems (approx. 20 laboratories per manufacturer/system). Participation also includes the in-house laboratories of the respective manufacturers. For each survey, we select eight different analytes from the clinical chemistry test menu of modern platforms (for the analytes covered up to now, see Table 1). The participants are requested to do the measurement in singlet under within-run conditions. The quality of assays and laboratories is assessed from four quality indicators at the peer group and "reference" level. The latter uses either the all manufacturer trimmed mean (AMTM) or reference method values as target: (i) the standard error of the estimate (% $S_{y/x}$) from linear regression analysis: if data are compared with the peer group mean, the $S_{y/x}$ is a measure for within-run imprecision; if compared with the reference target, it reflects the combined imprecision (both random and sample related effects); (ii) bias (%) at the mean concentration and the range limits (low and high concentration end); (iii) total error (%); (iv) the number of results observed outside the total error limits. These estimates are tested against a hierarchy of decision limits, i.e., limits that account for state-of-the-art performance, but also limits derived from biological variation data [9].

Patient percentile monitoring

We monitor the daily medians of the results for 20 commonly measured analytes in serum or plasma. All types and sizes of laboratories can participate. The laboratories calculate instrument-specific daily medians from outpatient results and send the data by e-mail to our database. Several vendors of laboratory information systems offer cost-free solutions for automatic calculation and electronic transfer. Alternative solutions are extraction of weekly/monthly data from the system and shipment in batch. Formats readable in our database are an e-mail-embedded table, Excel files, and text files (Supplementary Material, Table 1, that accompanies the article http://www.degruyter.com/view/j/cclm.2015.53.issue-8/cclm-2014-0959/cclm-2014-0959.xml?format=INT). Note that we do the mapping of the laboratories' mnemonics for the different analytes and units for expression of the medians. Via a user interface with authentication (access by user name and password) for secured authorization, the participating laboratory can plot for each analyte the course of the moving median. If a laboratory reports medians for different instruments, the moving medians (instrument-specific colored) are shown in the same plot. For interpretation, preliminary desirable limits for mid- to long-term bias are included. These are guided by biological variation and state-of-the-art performance (Supplementary Material,

Table 1 Overview of the analytes covered in the Empower project.

Analytes covered in the master comparisons and patient percentile monitoring	
Alanine aminotransferase[a]	Glucose[b]
Albumin[c]	Lactate dehydrogenase[a]
Alkaline phosphatase[a]	Magnesium[c]
Aspartate aminotransferase[a]	Phosphate[b]
Calcium[c]	Potassium[a]
Chloride[a]	Sodium[a]
Total cholesterol[b]	Total protein[c]
Creatinine[b]	Total triglycerides[b]
γ-Glutamyl transferase[a]	Uric acid (urate)[b]
Analytes only covered in the master comparisons	**Analytes only covered in the patient percentile monitoring**
HDL-cholesterol[b]	C-reactive protein
LDL-cholesterol[b]	Total bilirubin

Analytes covered in references [a][8], [b][6], and [c][5], respectively.

Table 2). The user application allows selection of (i) the number of consecutive medians (n=5, 8, 16) used for calculation of the moving median, (ii) time window, and (iii) exclusion of data from weekends. Each plot also shows the long-term median of the concerned individual laboratory as well as the peer group or all devices median (freely to select). Additional numerical information is provided on the long-term imprecision (the so-called robust CV, %) and the bias calculated in comparison to the peer group or all devices target as well as a "desirable" target. Currently, we use the medians of the reference intervals determined in the trueness-based "Nordic Reference Interval Project (NORIP)" as preliminary reference source for that target [10]. The user can download and print the plots. He has also access to his own entries in the database with the possibility to filter/sort according to analyte/date. This facilitates tracing back on which date graphical aberrant observations started. The graphical user interface can be accessed at https://www.thepercentiler.be/ (to see the demo version, log in with "demolab" as username and "demo1234" as password). A screenshot is given in the Supplementary Material, Figure 1.

Results

Status of the project

Results of the master comparison surveys conducted up to now are described elsewhere [4–6, 8]. The online Supplemental Figure 2 shows that in the most recent survey (2014), a total of 125 laboratories from 21 different countries (15 in Europe and Australia, Canada, Malaysia, South Korea, Singapore, and the USA) participated. The five main manufacturers also joined with their in-house laboratories (Abbott, Beckman, Ortho, Roche, and Siemens). In the patient percentile monitoring part, currently, 100 laboratories from 15 different countries (11 in Europe and Australia, India, Russia, and the USA) are participating with a total of 182 devices (Supplementary Material, Figure 2). Most of the test systems involved in the 2014 master comparison survey are also represented in percentile monitoring (Supplementary Material, Table 3). Table 1 shows that most analytes covered in the master comparisons (20 until now) are also addressed in patient percentile monitoring.

Test performance, comparability across manufacturers, and laboratory performance

As described elsewhere in detail, the design of the master comparisons with 20 single-donation commutable samples allows to assess different performance attributes of the examined assays and also individual laboratory performance [4–6]. Apart from some exceptions, assay peer group assessment showed a good intrinsic analytical quality in terms of within-run and combined imprecision and total error. It also demonstrated sufficient robustness for satisfactory performance in a daily laboratory context. However, there was room for improvement at higher and lower concentrations. Assessment at the reference level showed for several analytes good comparability between manufacturers/assays, e.g., for total protein, cholesterol, glucose, phosphate, uric acid [5, 6], whereas for others, considerable calibration differences were obvious, e.g., for albumin [5]. Particularly striking in this regard were the biases against the targets for enzymes set by the IFCC reference methods [8, 11–15]. Also, long-term assay drift/uncorrected biases for a single manufacturer were sometimes uncovered, e.g., magnesium, creatinine, low density lipoprotein (LDL) cholesterol, phosphate, uric acid, and chloride in [5, 6, 8]. Assessment against the reference

method or AMTM showed for most assays and analytes sufficient analytical specificity, but for others, vulnerability to sample-related effects, e.g., high density lipoprotein (HDL) and LDL cholesterol in [6]. The bias limits used for assessment demonstrated that for certain analytes, the state of the art is such that most assays, apart from some, can meet the desirable biological variation bias specifications (e.g., for total protein, phosphate, triglycerides, uric acid, alkaline phosphatase, potassium [5, 6, 8]). For some biologically more tightly regulated analytes, the biologically inferred limits are not feasible, e.g., glucose, cholesterol, chloride [6], or would require improvement of lot-to-lot consistency, e.g., calcium [5]. In contrast, sodium assays showed exceptionally well performing, almost within the tight biological bias limit [8]. Assessment of the laboratory performance strikingly showed that sometimes large between-laboratory differences (>30%) occurred for all analytes [6, 8]. These discrepancies could partly be ascribed to the biases in the used assays but also likely point to severe laboratory effects on performance of assays in daily practice.

Similar observations were made from the patient percentile monitoring data. For example, the median values matched the aforementioned calibration differences revealed for γ-glutamyl transferase and chloride in the 2014 master comparison survey (Figure 1) [8]. Indeed the γ-glutamyl transferase moving median values ranged from approximately 20 to approximately 32 U/L, those for chloride from approximately 101 to approximately 105 mmol/L.

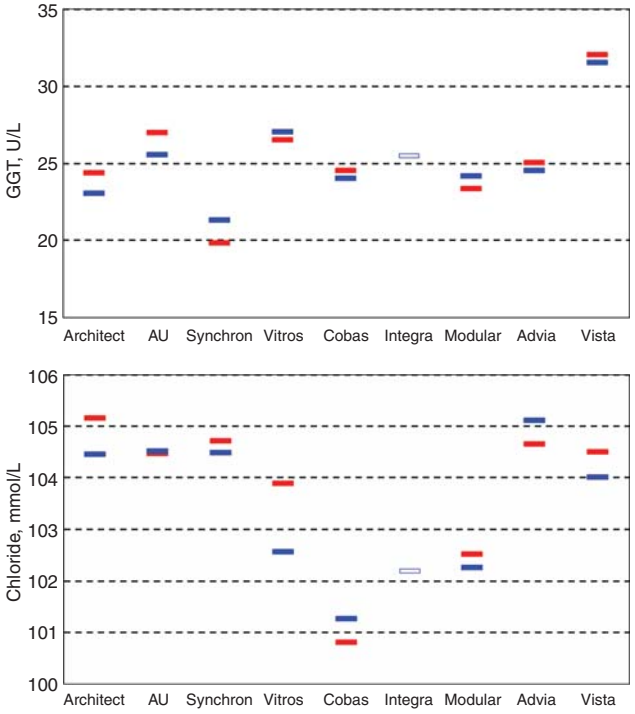

Figure 1 Illustration of the match between the peer group means (red rectangles) in the 2014 survey of the master comparisons and the median values (blue rectangles) in patient percentile monitoring for γ-glutamyl transferase (GGT) and chloride.

Stability of laboratory/test performance

First, results from patient percentile monitoring show that laboratories with high daily throughput and/or low variation in patient population typically perform with low variation and mostly good concordance between the different instruments (Supplementary Material, Figure 3A). Other laboratories have a higher long-term variation in performance. If this is due to a lower throughput or higher population variation (typical for laboratories operating in a medium-sized hospital), the variation can partly be reduced by selecting a higher n for calculation of the moving median (Supplementary Material, Figure 3B and C). Other observations are about drifts or shifts, transient to long-term bias, e.g., between different instruments used in a laboratory, of one particular instrument compared with the others, or of the laboratory compared with its peer. Interestingly, shifts or drifts sometimes apply for several laboratories belonging to the same peer, which confirms that they are caused by a major manufacturer event, e.g., a reagent or calibrator lot change (Figure 2A and B). In other cases, laboratories can relate the observed instability to a calibration event (example shown in Figure 2C). Although certain observations can rather easily be explained, longer observation times and more solid peer groups are needed for a systematic investigation of the root causes.

Discussion

The Empower project is an integrated quality assurance tool for laboratories and manufacturers. Its unique design based on real patient results allows to assess/demonstrate quality aspects without being confounded by commutability issues [16, 17]. It facilitates remediating actions because it reveals major bias components/sources, such as the manufacturer (assay), laboratory, instrument, the reagent/calibrator lot, and recalibration by the laboratory itself (Figure 3).

The focus of the master comparisons, which are conducted across assays and laboratories, is on how well the intrinsic analytical quality of assays on release by the respective manufacturers is reproduced by the end

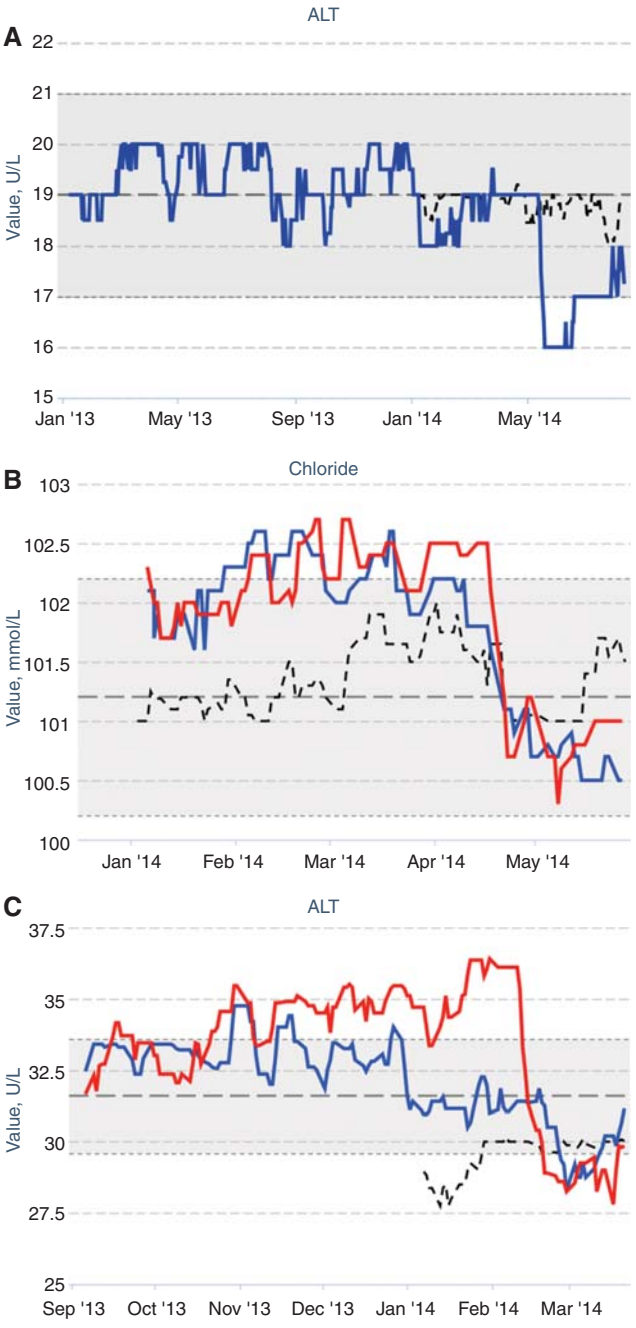

Figure 2 Significant test instability for alanine aminotransferase (ALT) and chloride due to a confirmed reagent lot change (A), a calibrator lot change (B), and a laboratory calibration event (C).
In (A) and (B), it is illustrated how lot changes can disturb the stable performance. The long broken gray line represents the median calculated from all daily medians provided by the laboratory to which the graph applies. In (C), the moving median for one of the instruments (red colored full line) started to drift around the 20th of December 2013, and on 12th February 2014, both instruments (also the blue one) were recalibrated by the laboratory, which caused in both a shift. The shifts moved the medians outside the stability zone (shaded area between short broken gray lines). The black short broken line represents the peer group moving medians in (A), (B), and (C).

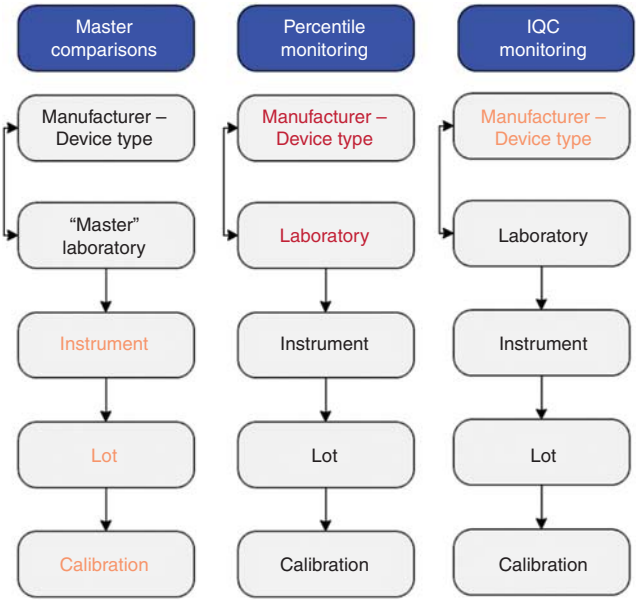

Figure 3 Assessment (and control) of bias components/sources. Components in black can be assessed by the above standing pillar, those in orange cannot; those in red probably also can in high-throughput laboratories that mainly work with general practitioners (samples almost exclusively from outpatients).

users under "field" conditions all over the world. Target setting is based on reference method measurements or the AMTM. These targets allow to assess either the real traceability (standardization status) or the comparability between assays and laboratories. The information on traceability is of utmost use for the discipline of clinical chemistry to investigate the extent of implementation of standardization efforts. Note in this regard the striking example of the bias observed for enzyme assays. For the individual laboratory, it is first-hand information that can help in decisions on the acquisition of new instruments. As such, the master comparisons provide the participating laboratories with a reliable calibration fix-point of their own performance within the peer group and of the latter compared with other peers. Naturally, this is only a point estimate in time that should continuously be monitored. This is where patient percentile (and IQC) monitoring comes into play. Indeed, the stability of the peer group calibration fix-points can be appreciated from concordant medians from outpatient results (Figure 1). In addition, laboratories can use their medians as a tool to monitor the mid- to long-term stability of their own calibration status, again in comparison to their peer, and/or to uncover shifts/drifts and the sources thereof [18]. Of course, this requires that the moving medians in time truly reflect the analytical variation, without being confounded by other sources of variation. In a pilot study, we showed that by

working with medians from outpatients and omitting medians from weekends and holidays (days with lower throughput and/or altered ratios of inpatients to outpatients), the effect of patient population variability can be suppressed. We inferred this from a congruent course in time of the moving medians and mean of daily IQC data [18]. Meanwhile, it is our experience that in high-throughput laboratories mainly serving outpatients, the moving medians can be calculated from a low number of daily medians (n=5). This is the ideal number for detection of analytical instabilities (shifts, drifts). In contrast, for laboratories in a hospital context, a higher n is required to partly compensate for the effect of a more variable patient population and lower throughput. We offer in the user interface n=8 or 16, but the latter is the limit needed to prevent too much smoothing and loss of resolution. Another asset of the percentile monitoring design is that it shows the instrument-specific stability in one plot. This allows the laboratories to monitor the interchangeability of results among different instruments and/or detect the occurrence of instrument-specific special events.

Notwithstanding the above potential of the percentile monitoring tool, we recommend the users to do the interpretation with caution. Indeed, certain influential factors may explain aberrant or more variable medians. We learned, for example, that in hospital laboratories, dialysis or oncology patients are often registered as outpatient and that their samples are preferentially measured on one instrument. Note, however, that by closely working with our participants, we can share our experience to enable more critical interpretation. We also recommend sample exchange between partner laboratories belonging to the same peer group and preferably participating in patient percentile monitoring because this may be very helpful to exclude or confirm observed laboratory biases.

We want to emphasize that monitoring of patient medians is not a substitute for daily IQC. We advocate it as a complementary observation tool from patient data that can cover much longer observation times.

A fundamental question in all parts of the Empower project is whether the observed differences in quality of performance or instability are to be considered significant. This points to the importance of performance specifications for meaningful conclusions [19–22]. In the absence of a consensus on this topic, we use preliminary limits that are guided by biological variation [9], and also by state-of-the-art performance (for the master comparisons, we refer to [6]; for the patient percentile monitoring, see online Supplemental Table 2). This means that for tightly regulated analytes, we expand the limits based on biological variation to account for the current quality offered by manufacturers. Note that for patient percentile monitoring, we express the limits for allowable bias in absolute terms (tailored to the used SI units). The reason is that this allows us to show them in the user interface as so-called stability limits that should not be exceeded by longer than 1 week. See, for example, the shaded zone between 17 and 19 U/L (median±2 U/L) in Figure 2A for alanine aminotransferase. Interestingly, we found the patient percentile monitoring an excellent tool to test how realistic our quality goals are, e.g., the stability limit of 1 mmol/L for sodium [23]. For other analytes with very high biological variation, such as C-reactive protein, we set a general upper limit of approximately 10%.

Another important question is which targets to use. For the master comparisons part, this is discussed elsewhere [6]. In the percentile monitoring part, we compare the medians in first instance not only with the peer group medians but also with a reference median. We use the median from the NORIP reference intervals, which is, to the best of our knowledge, the only source that claims to be “trueness-based” [10]. The reliability is high for analytes such as sodium and calcium, but the information for some enzymes has to be interpreted critically. There have been changes in the IFCC-recommended methods, and it is known that these are either not carefully or uniformly adopted by manufacturers. Therefore, we still consider NORIP as a preliminary reference source and will follow up, e.g., by cross-comparison with the reference interval information from manufacturers and new projects.

Of course, the utility of our project has to be improved on a continuous basis. For example, we aim at a platform that stimulates the dialogue on a basis of trust between the participant laboratories and manufacturers. We work on this by establishing close contacts with both parties. We also plan to develop a new tool that investigates the effect of analytical (in)stability on a surrogate medical outcome, such as the frequency of “flagged results” [18]. Together with realistic quality goals that result in meaningful conclusions, this tool might be an excellent basis to strengthen the physician/laboratory interface by more transparent communication on performance. The Empower database potentially can become a source for “big data mining” with utility for studies that relate the outcome of therapeutic strategies to median values in patient cohorts (e.g., the Dialysis Outcomes and Practice Patterns Study) [24]. From the perspective that the project’s general emphasis is on interchangeability of laboratory results, it can potentially also contribute to modern clinical needs such as the definition of common reference intervals or clinical decision limits, implementation of

electronic health records, and development of evidence-based clinical practice guidelines for application of consistent standards of medical care.

Conclusions

The Empower project provides evidence on the intrinsic quality of assays and how this quality is sustained under field conditions. It also demonstrates how well assays and laboratories compare and how stable they perform. In addition, it enables to uncover all major bias components/sources. The major asset of the project is that it works with data generated from real patient samples and can be linked to observations in daily IQC practice. From this perspective, we believe it is a new integrated tool for modern quality management of benefit to all stakeholders with interest in reliable laboratory data. It can help the discipline of clinical chemistry to derive realistic quality specifications and can strengthen the laboratory/manufacturer dialogue and laboratory/physician interface. Ultimately, if the evidence provided by the project is translated into action by laboratories and manufacturers, it can contribute to a yet to be established translational laboratory medicine and better patient care.

Acknowledgments: The authors are indebted to the laboratories and diagnostic manufacturers who showed their interest for the Empower project, either by participating in one or more surveys of the master comparisons and/or joining the patient percentile monitoring initiative.
Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Financial support: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Miller WG. The role of proficiency testing in achieving standardization and harmonization between laboratories. Clin Biochem 2009;42:232–5.
2. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. Clin Chem 2011;57:76–83.
3. Sciacovelli L, Secchiero S, Zardo L, Zaninotto M, Plebani M. External quality assessment: an effective tool for clinical governance in laboratory medicine. Clin Chem Lab Med 2006;44:740–9.
4. Stepman HC, Stöckl D, Acheme R, Sesini S, Mazziotta D, Thienpont LM. Status of serum-calcium and -albumin measurement in Argentina assessed in 300 representative laboratories with 20 fresh frozen single donation sera. Clin Chem Lab Med 2011;49:1829–36.
5. Van Houcke SK, Rustad P, Stepman HC, Kristensen GB, Stöckl D, Røraas TH, et al. Calcium, magnesium, albumin, and total protein measurement in serum as assessed with 20 fresh-frozen single-donation sera. Clin Chem 2012;58:1597–9.
6. Stepman HC, Tiikkainen U, Stöckl D, Vesper HW, Edwards SH, Laitinen H, et al. Measurements for 8 common analytes in native sera identifies inadequate standardization among 6 routine laboratory assays. Clin Chem 2014;60:855–63.
7. CLSI. Preparation and validation of commutable frozen human serum pools as secondary reference materials for cholesterol measurement procedures; approved guideline. CLSI document C37-A. Wayne, PA: Clinical and Laboratory Standards Institute, 1999.
8. STT-Consulting – Empower – Master Comparison 2014. http://www.stt-consulting.com/news.php?rubriek=8/. Accessed September 2014.
9. Westgard QC. Biological variation database, and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update. http://www.westgard.com/biodatabase-2014-update.htm. Accessed September 2014.
10. Nordic Reference Interval Project (NORIP). http://pweb.furst.no/norip/. Accessed September 2014.
11. Schumann G, Bonora R, Ceriotti F, Clerc-Renaud P, Ferrero CA, Férard G, et al. IFCC Primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37°C. Part 3. Reference procedure for the measurement of catalytic concentration of lactate dehydrogenase. Clin Chem Lab Med 2002;40:643–8.
12. Schumann G, Bonora R, Ceriotti F, Férard G, Franck PF, Gella F-J, et al. IFCC Primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37°C. Part 4. Reference procedure for the measurement of catalytic concentration of alanine aminotransferase. Clin Chem Lab Med 2002;40:718–24.
13. Schumann G, Bonora R, Ceriotti F, Férard G, Ferrero CA, Franck PF, et al. IFCC Primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37°C. Part 5. Reference procedure for the measurement of catalytic concentration of aspartate aminotransferase. Clin Chem Lab Med 2002;40:725–33.
14. Schumann G, Bonora R, Ceriotti F, Férard G, Ferrero CA, Franck PF, et al. IFCC Primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37°C. Part 6. Reference procedure for the measurement of catalytic concentration of γ-glutamyltransferase. Clin Chem Lab Med 2002;40:734–8.
15. Schumann G, Klauke R, Canalias F, Bossert-Reuther S, Franck PF, Gella F-J, et al. IFCC Primary Reference procedures for the measurement of catalytic activity concentrations of enzymes at 37°C Part 9. Reference procedure for the measurement of catalytic concentration of alkaline phosphatase. Clin Chem Lab Med 2011;49:1439–46.

16. Stöckl D, Thienpont LM. The combined-target approach: a way out of the proficiency testing dilemma. Arch Pathol Lab Med 1994;118:775–6.
17. Horowitz GL. Assessing accuracy on the front lines: a pragmatic approach for single-donor proficiency testing. Clin Chem 2014;60:806–8.
18. Van Houcke SK, Stepman HC, Thienpont LM, Fiers T, Stove V, Couck P, et al. Long-term stability of laboratory tests and practical implications for quality management. Clin Chem Lab Med 2013;51:1227–31.
19. Tonks DB. A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. Clin Chem 1963;9:217–33.
20. Kallner A, McQueen M, Heuck C. The Stockholm Consensus Conference on quality specifications in laboratory medicine, 25–26 April 1999. Scand J Clin Lab Invest 1999;59:475–6.
21. IFCC Working group on allowable errors for traceable results (WG-AETR). http://www.ifcc.org/ifcc-scientific-division/sd-working-groups/allowable-errors-for-traceable-results-wg-aetr/. Accessed September 2014.
22. 1st EFLM Strategic Conference – defining analytical performance goals – 15 years after the Stockholm Conference. http://www.efcclm.eu/files/efcc/Leaflet%20EFLM%20strategic%20conference.pdf. Accessed September 2014.
23. Stepman HC, Stöckl D, Stove V, Fiers T, Couck P, Gorus F, Thienpont LM. Long-term stability of clinical laboratory data – sodium as benchmark. Clin Chem 2011;57:1616–7.
24. Arbor Research Collaborative for Health. Dialysis Outcomes and Practice Patterns Study. http://www.dopps.org. Accessed September 2014.

Supplemental Material: The online version of this article (DOI: 10.1515/cclm-2014-0959) offers supplementary material, available to authorized users.